

[19] 中华人民共和国国家知识产权局

[51] Int. Cl<sup>7</sup>

G10L 15/10

G10L 13/00



# [12] 发明专利申请公开说明书

[21] 申请号 01808021.9

[43] 公开日 2003 年 6 月 11 日

[11] 公开号 CN 1423809A

[22] 申请日 2001.4.10 [21] 申请号 01808021.9

[30] 优先权

[32] 2000. 4. 14 [33] JP [31] 114262/2000

[86] 国际申请 PCT/JP01/03079 2001.4.10

[87] 国际公布 WO01/80222 日 2001.10.25

[85] 进入国家阶段日期 2002.10.14

[71] 申请人 酒井康江

地址 日本埼玉县

[72] 发明人 小柳裕喜生

[74] 专利代理机构 中科专利商标代理有限责任公  
司

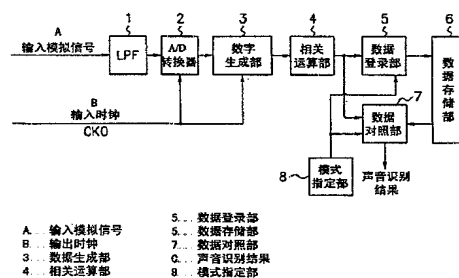
代理人 李香兰

权利要求书 4 页 说明书 24 页 附图 12 页

[54] 发明名称 声音识别方法及装置、声音合成方法  
及装置及记录媒体

## [57] 摘要

对输入声音信号进行微分，而检测出以该微分  
值所符合预定条件的点为标本的点，且在取得检测  
后各标本点的离散振幅数据与表示各标本点间的时  
间间隔的定时数据后，利用上述振幅数据及定时数  
据来产生相关数据，通过对照该产生的相关数据与  
针对各种声音事先予以同样产生而储存的相关数据  
来识别输入声音。



1. 一种声音识别方法，其特征在于：

5       对输入声音信号进行微分，而检测出以该微分值所符合预定条件的点为标本的点，且在取得检测后各标本点的离散振幅数据与表示各标本点间的时间间隔的定时数据后，利用上述振幅数据及定时数据产生相关数据，通过对照该产生的相关数据与针对各种声音事先予以同样产生而储存的相关数据识别输入声音。

10       2. 根据权利要求 1 所述的声音识别方法，其特征在于：以上述输入声音信号的微分绝对值为形成预定值以下的点的时间间隔使其标本化。

3. 根据权利要求 1 所述的声音识别方法，其特征在于：以上述输入声音信号的微分绝对值为形成极小的点的时间间隔使其标本化。

15       4. 根据权利要求 1 所述的声音识别方法，其特征在于：以上述输入声音信号的微分值的极性所变化的点的时间间隔使其标本化。

5. 根据权利要求 1 所述的声音识别方法，其特征在于：上述相关数据为连续的标本点的振幅数据彼此的比率及连续的标本点间的定时数据彼此的比率。

20       6. 根据权利要求 1 所述的声音识别方法，其特征在于：进行使上述相关数据的下位数位化整的处理。

7. 根据权利要求 1 所述的声音识别方法，其特征在于：过取样上述输入声音信号，以该过取样的数据的微分值所符合预定条件的点的时间间隔来使其标本化。

25       8. 根据权利要求 7 所述的声音识别方法，其特征在于：通过过取样与移动平均运算或卷积运算来合成对应于使上述输入声音信号数字化而取得的  $n$  个离散数据的值的基本波形的数字数据，由此在对上述离散数据求得数位的插入值后，以上述求得的数位插入值的微分值所符合预定条件的点的时间间隔使其标本化。

30       9. 一种声音识别装置，其特征在于：具备  
一 A / D 转换装置；该 A / D 转换装置是供以对输入声音信号进行 A

/D 转换; 及

一微分单元; 该微分单元是供以对上述 A/D 转换装置所输出的数字数据进行微分; 及

5 一数据产生单元; 该数据产生单元是供以对输入声音信号进行微分, 而检测出以该微分值所符合预定条件的点为标本的点, 产生检测后各标本点的离散振幅数据与表示各标本点间的时间间隔的定时数据; 及

一相关运算单元; 该相关运算单元是供以利用上述数据产生单元所产生的上述振幅数据及定时数据来产生相关数据; 及

10 一数据对照单元; 该数据对照单元是供以对照上述相关运算单元所产生的相关数据与针对各种声音事先予以同样产生而储存于记录媒体的相关数据而识别输入声音。

10. 根据权利要求 9 所述的的声音识别装置, 其特征在于: 上述数据产生单元是以上述 A/D 转换装置所输出的数字数据的微分绝对值为形成预定值以下的点的时间间隔来使其标本化。

15 11. 根据权利要求 9 所述的的声音识别装置, 其特征在于: 上述数据产生单元是以上述 A/D 转换装置所输出的数字数据的微分绝对值为形成极小的点的时间间隔来使其标本化。

12. 根据权利要求 9 所述的的声音识别装置, 其特征在于: 上述数据产生单元是以上述 A/D 转换装置所输出的数字数据的微分值的极性所变化的  
20 点的时间间隔来使其标本化。

13. 根据权利要求 9 所述的的声音识别装置, 其特征在于: 上述相关运算单元是供以求得连续的标本点的振幅数据彼此的比率及连续的标本点间的定时数据彼此的比率来作为上述相关数据。

14. 根据权利要求 9 所述的的声音识别装置, 其特征在于: 上述相关运  
25 算单元是在于进行使上述相关数据的下位数位化整的处理。

15. 根据权利要求 9 所述的的声音识别装置, 其特征在于: 具备

利用偶数倍频率的时钟来对上述 A/D 转换装置所输出的数字数据进行过取样的过取样单元;

上述数据产生单元是以上述过取样的数据的微分值为符合预定条件的  
30 点的时间间隔使其标本化。

16. 根据权利要求 15 所述的声​​音识别装置，其特征​​在于：上述过取样单元是通过过取样与移动平均运算或卷积运算来合成对应于上述 A / D 转换装置所输入的 n 个离散数据的值的基本波形的数字数据，以对上述离散数据求得数位的插入值。

5        17. 一种声音合成方法，其特征​​在于：

      使声音以外的数据与针对对应于该数据的声音信号而事先产生的上述声音信号的微分值为符合预定条件的标本点的振幅数据及表示各标本点间的时间间隔的定时数据的组合彼此相联系，当所期望的数据被指定时，利用与该被指定的数据相联系的上述振幅数据及定时数据的组合，求得插入  
10        具有根据上述定时数据而表示的时间间隔的上述振幅数据间的插入数据，由此来合成声音。

      18. 根据权利要求 17 所述的声​​音合成方法，其特征​​在于：利用由连续的 2 个标本点的 2 个振幅数据及其间的定时数据所取得的有限标本化函数来求得插入上述 2 个振幅数据间的插入数据。

15        19. 一种声音合成装置，其特征​​在于：具备

      一存储单元；该存储单元是使针对对应于声音以外的数据的声音信号而事先产生的上述声音信号的微分值为符合预定条件的标本点的振幅数据及表示各标本点间的时间间隔的定时数据的，组合彼此与上述声音以外的数据相联系，然后再予以存储起；及

20        一插入处理单元；该插入处理单元是在所期望的数据被指定时，利用与该被指定的数据相联系的上述振幅数据及定时数据的组合来求得插入具有根据上述定时数据而表示的时间间隔的上述振幅数据间的插入数据；及

      一 D / A 转换装置；该 D / A 转换装置是供以对插入处理单元所求得的插入数据进行 D / A 转换。

25        20. 根据权利要求 19 所述的声​​音合成装置，其特征​​在于：具备一定时控制单元，该定时控制单元是根据表示由上述存储单元所读出的各标本点间的时间间隔的定时数据，以能够在上述各标本点间的时间间隔依次读出各标本点的振幅数据的方式控制时间；

      上述插入处理单元是利用根据上述定时控制单元所读出的连续 2 个标本点的 2 个振幅数据及其间的定时数据，求得插入上述 2 个振幅数据间的  
30

插入数据。

21. 根据权利要求 20 所述的声合成装置, 其特征在于: 上述插入处理单元是利用由上述, 连续的 2 个标本点的 2 个振幅数据及其间的定时数据所取得的有限标本化函数, 求得插入上述 2 个振幅数据间的插入数据。

22. 一种电脑可读取的记录媒体, 其特征是在于记录有一种程序, 该程序以便使权利要求 1 所记载的声音识别方法的处理程序执行于电脑。

23. 一种电脑可读取的记录媒体, 其特征在于记录有一种程序, 该程序以便使权利要求 9 所记载的各单元的功能作用于电脑。

24. 一种电脑可读取的记录媒体, 其特征在于记录有一种程序, 该程序以便使权利要求 17 所记载的声音合成方法的处理程序执行于电脑。

25. 一种电脑可读取的记录媒体, 其特征在于记录有一种程序, 该程序以便使权利要求 19 所记载的各单元的功能作用于电脑。

## 声音识别方法及装置、声音合成方法及装置及记录媒体

5

## 技术领域

本发明是有关声音识别方法及装置，声音合成方法及装置，以及储存有供以利用软件来实现这些方法及装置的功能的程序的记录媒体，特别是有关分辨人等的声音的声音识别技术，及由数据来制作声音的声音合成技术。

10

## 背景技术

近年来，通过电脑来识别人所发出的声音，而自动进行处理的所谓声音识别的声音处理界面颇受注视。就声音识别技术而言，例如有识别数字及单字等的单字声音识别，或者是理解意义及内容的连续声音识别等。并且还包含识别出所发出的声音是谁的话的识别技术。

15

尤其是最近声音识别的技术有由单字声音识别转移到连续声音识别或以自然会话为对象的趋势。甚至，利用特定的语汇来作为声音指令，而根据声音来操作计算机或携带型终端机等方法的研究也正进行著。

20

另一方面，由原文等的数据来合成声音而输出的声音合成技术的开发也正被进行。例如，就原文声音合成的技术而言，是在于解析由汉字或英文字等种种的文字所构成的语言的原文数据，根据事先设定的规则来赋予重音及语调，而由此来合成声音。

但，就目前的声音识别技术而言，并非是何发出的声音皆可识别，其识别率有限度。即使是同样的单字，其声音也会随着出声者的不同而有所差异，即识别精度会依说话者而产生差异。而且，一旦识别对象的语汇数或话者增加，其声音识别的难易度会更为增加。

25

另一方面，就目前的声音合成技术而言，所产生的合成音尚无法脱离机械音域，如何取得更接近人的声音，即如何达成高质量化将成为课题。

另外，由于目前的声音识别技术及声音合成技术皆须复杂的处理，因

30

此为了谋求处理的高速化等，而必须追求处理的简略化。

### 发明内容

本发明是鉴于上述实情而研发的，即为了谋求声音识别率的提高，及合成音的高质量化，以及处理的简略化等，其目的在于提供一种前所未有的全新声音识别方式及声音合成方式。

本发明的声音识别方法，其特征在于：

对输入声音信号进行微分，而检测出以该微分值所符合预定条件的点为标本的点，且在取得检测后各标本点的离散振幅数据与表示各标本点间的时间间隔的定时数据后，利用上述振幅数据及定时数据来产生相关数据，通过对照该产生的相关数据与针对各种声音事先予以同样产生而储存的相关数据来识别输入声音。

在此，也可进行使上述相关数据的下位数位化整的处理。

另外，也可过取样上述输入声音信号，以该过取样的数据的微分值所符合预定条件的点的时间间隔来使其标本化。

另外，本发明的声音识别装置，特征在于具备：

一 A / D 转换装置；该 A / D 转换装置是供以对输入声音信号进行 A / D 转换；及

一微分单元；该微分单元是供以对上述 A / D 转换装置所输出的数字数据进行微分；及

一数据产生单元；该数据产生单元是供以对输入声音信号进行微分，而检测出以该微分值所符合预定条件的点为标本的点，产生检测后各标本点的离散振幅数据与表示各标本点间的时间间隔的定时数据；及

一相关运算单元；该相关运算单元是供以利用上述数据产生单元所产生的上述振幅数据及定时数据来产生相关数据；及

一数据对照单元；该数据对照单元是供以对照上述相关运算单元所产生的相关数据与针对各种声音事先予以同样产生而储存于记录媒体的相关数据，而来识别输入声音。

在此，上述相关运算单元是在于进行使上述相关数据的下位数位化整的处理。

另外，也可具备：利用偶数倍频率的时钟来对上述 A / D 转换装置所输出的数字数据进行过取样的过取样单元，以上述过取样的数据的微分值为符合预定条件的点的时间间隔来使其标本化。

另外，本发明的声音合成方法，其特征在于：

- 5       使声音以外的数据与针对对应于该数据的声音信号而事先产生的上述声音信号的微分值为符合预定条件的标本点的振幅数据及表示各标本点间的时间间隔的定时数据的组合彼此相联系，当所期望的数据被指定时，利用与该被指定的数据相联系的上述振幅数据及定时数据的组合来求得插入具有根据上述定时数据而表示的时间间隔的上述振幅数据间的插入数据，  
10       以合成声音。

另外，本发明的声音合成装置的特征在于具备：

- 一存储单元；该存储单元是使针对对应于声音以外的数据的声音信号而事先产生的上述声音信号的微分值为符合预定条件的标本点的振幅数据及表示各标本点间的时间间隔的定时数据的组合彼此与上述声音以外的数  
15       据相联系，然后再予以存储；及

一插入处理单元；该插入处理单元是在所期望的数据被指定时，利用与该被指定的数据相联系的上述振幅数据及定时数据的组合来求得插入具有根据上述定时数据而表示的时间间隔的上述振幅数据间的插入数据；及

- 一 D / A 转换装置；该 D / A 转换装置是供以对插入处理单元所求得的插入数据进行 D / A 转换。  
20

另外，本发明的电脑可读取的记录媒体的特征是在于记录程序，该程序是供以使本发明的声音识别方法，或声音合成方法的处理程序执行于电脑。

- 另外，本发明的电脑可读取的记录媒体的特征是在于记录程序，该程序是供以便本发明的各单元的功能作用于电脑。  
25

由于本发明是根据上述技术单元所构成，因此可提供一种利用预定标本点的振幅数据与定时数据之间所未有的全新声音识别方法及声音合成方法。

- 在此，并非是原封不动地利用振幅数据及定时数据，而是能够利用其相关数据来提高声音的识别率。并且，可对相关数据的下位数位进行化整  
30



处理或声音信号的过取样处理，而使声音的识别率能够更为提高。

另外，可利用由实际的声音信号所构成的振幅数据及定时数据来合成声音，由此只要存储少数的数据便能够合成更接近真实的声音。由于是利用声音信号的过取样处理而产生的振幅数据及定时数据来进行声音的合成；因此能够更为提高合成音的质量。

如上所述，若利用本发明，则可谋求声音识别率的提高，及合成音的高质量化，以及处理的简略化。

### 附图说明

- 10 图 1 是表示实施例 1 的声音识别装置的构成例方块图。  
图 2 是用以说明本实施例的声音识别的原理图。  
图 3 是表示数据产生部的构成例方块图。  
图 4 是表示图 3 的微分器的构成例。  
图 5 是表示通过进行 2 重微分来检测出标本点的构成例方块图。  
15 图 6 是表示实施例 1 的声音合成装置的构成例方块图。  
图 7 是用以说明本实施例的声音合成的原理图。  
图 8 是取出图 7 的时间 T1—T2 的区间来表示本实施例的插入原理图。  
图 9 是表示标本化函数的例子。  
图 10 是表示供以声音合成的插入动作说明图。  
20 图 11 是用以说明数据插入处理的具体例的插入运算式。  
图 12 是表示实施例 2 的声音识别装置的构成例方块图。  
图 13 是表示利用于实施例 2 的数位基本波形图。  
图 14 是用以说明实施例 2 的过取样及卷积运算的动作例图。  
图 15 是表示从实施例 2 的数位基本波形产生的函数。  
25 图 16 是表示图 12 的过取样电路的构成例图。

### 具体实施方式

#### 实施例 1

以下，根据附图说明本发明的一实施例。

- 30 图 1 是表示本实施例的声音识别装置的构成例方块图。如图 1 所示，

本实施例的声音识别装置是由：低通滤波器（LPF）1，A/D转换器2，数据产生部3，相关运算部4，数据登记部5，数据存储体6，数据对照部7，及模式指定部8所构成。

在图1中，模拟输入信号是从麦克风（未图示）等来输入人等所发出的声音。该模拟输入信号为了能够容易进行后述的标本点的检测，而在通过LPF1来去除杂波后，利用A/D转换器2来转换成数字数据。此刻，A/D转换器2是根据预定频率（例如，44.1KHz）的输入时钟CK0来执行A/D转换处理，例如将模拟输入信号转换成16位的数字数据。

并且，通过该A/D转换器2而被数字化的声音数据会被输入数据产生部3。而且，数据产生部3会对A/D转换器2所供给的数字数据进行微分，按照微分结果来检测出后述的标本点。然后，求得表示各检测点的数字数据的振幅的振幅数据，及表示各标本点间的时间间隔的定时数据（时钟CK0的数目），而输出。

图2是用以说明本实施例的声音识别的原理图。并且，被输入数据产生部3的数据虽是将模拟输入信号予以A/D转换后的数字数据，但在图2中为了便于说明，而将A/D转换器2所输出的数字数据的波形显示成模拟。而且，该图2中所示的数值为说明用的数值，非实际的数值。

在本实施例中，是例如由A/D转换器2所输出的数字数据101中检测出微分绝对值（信号的倾斜）为含“0”的预定值以下的点（将此称为标本点）102a—102f。然后，求得表示各标本点102a—102f的阵振幅的数字数据值，及表示各标本点102a—102f出现的时间间隔的定时数据值，且输出该振幅数据值与定时数据值的组合。

就图2的例子而言，各标本点102a—102f的数位的振幅数据值为“7，3，9，1，6，3”。另外，表示各标本点102a—102f出现的时间T1—T2间，T2—T3间，T3—T4间，T4—T5间，T5—T6间的各时间间隔为“5，7，3，3，3”。另外，在此定时数据所示的数字是根据某取样频率而示的时钟CK0的数目。

由于在时间T1的时间点可取得标本点102a的振幅数据值“7”，及表示来自在此之前标本点（未图示）所被检测出的时间的时间间隔的定时数据值（未图示），因此可将这些数据值的组合当作时间T1的数据而输

出。

其次，由于在标本点 102b 所被检测出的时间 T2 的时间点可取得表示来自在此之前标本点 102a 所被检测出的时间 T2 的时间间隔的定时数据

5 (5, 3) 当作时间 T2 的数据而输出。

其次，由于在标本点 102c 所被检测出的时间 T3 的时间点可取得表示来自在此之前标本点 102b 所被检测出的时间 T2 的时间间隔的定时数据

值“7”，及标本点 102c 的振幅数据值“9”，因此可将这些数据值的组合

(7, 9) 当作时间 T3 的数据而输出。

10 以下同样的可分别将表示 T3—T4 间，T4—T5 间，T5—T6 间的时间间隔的定时数据与在时间 T4, T5, T6 所被检测出的各标本点 102b, 102e, 102f 的振幅数据值的组合 (3, 1), (3, 6), (3, 3) 当作时间 T4, T5, T6 的数据而输出。

图 3 是表示数据产生部 3 的一构成例的方块图。在图 3 中，微分器 301

15 会在每个取样点来微分根据 A / D 转换器 2 而输入的数字数据，且取其绝对值来输出至标本点检测部 302。另外，标本点检测部 302 会根据上述微分器 301 所微分的结果来检测出数字数据的微分绝对值为形成预定值以下的标本点。

图 4 是表示上述微分器 301 的一构成例。如图 4 所示，本实施例的微

20 分器 301 是由运算连续 2 个取样点的数据间的差分绝对值的差分绝对值电路所构成。

在图 4 中，差分器 31, 32 是分别运算节点 a, b 所输入的连续 2 个取样点的数据差分。即，差分器 31 是运算差分  $a-b$ ，差分器 32 是运算差分  $b-a$ ，并且分别将其结果输出至 OR 电路 33, 34。这些差分器 31, 32

25 在所被运算的差分值为形成负值时，除了差分值以外，还输出作为借位用的“1”值。

上述 OR 电路 33 是取上述微分器 301 所运算的差分值与借位输出的逻辑和，并将其结果输出至 AND 电路 35。另外，另一 OR 电路 34 是取上述微分器 302 所运算的差分值与借位输出的逻辑和，并将其结果输出

30 至，AND 电路 35。另外，AND 电路 35 是取来自上述 2 个 OR 电路 33,

34 的逻辑乘积，并将其结果输出至节点 c。另外，将上述差分器 31 的借位输出予以输出至节点 d，将上述差分器 32 所运算的差分值予以输出至节点 e。

由此，在节点 c 会被输出连续 2 个取样点的数据的差分绝对值  $|a-b|$ ，另外，当节点 b 的数据值比节点 a 的数据值还要大时，在节点 d 会被输出“1”的值，另外，在节点 e 会被输出节点 a, b 的数据间的差分值  $b-a$ 。

在图 4 中，为了便于说明，而只将节点 a, b, c, e 的数据线予以显示出 1 位份，但实际上只具备数据的位数。

图 3 的定时产生部 303 是在于计数从 1 个标本点被检测出开始到下个标本点被检测出为止所供给的时钟 CK0 的数目，并予以作为时间数据输出的同时，输出表示各标本点的检测点时间的定时时钟。

另外，振幅产生部 304 是根据定时产生部 303 所输出的定时时钟来只取出该当标本点位置的数字数据，而作为振幅数据输出。如此一来，根据振幅产生部 304 所产生的各标本点的振幅数据，及表示根据定时产生部 303 所产生的各标本点间的时间间隔的定时数据的组合会被输出至图 1 的相关运算部 4。

另外，相关运算部 4 会针对上述数据产生部 3 所输出的各振幅数据及各定时数据的相关进行运算。就此相关运算的方法而言，有各种方式。在此为运算数据产生部 3 所输出的各振幅数据彼此的比例及各定时数据彼此的比例。例如，若取得振幅数据  $d1, d2, d3, d4\cdots$  及定时数据  $t1, t2, t3, t4\cdots$ ，则相关运算部 4 可计算出式 (1a), (1b) 的比例。

$$d2 / d1, d3 / d2, d4 / d3\cdots \quad (1a)$$

$$t2 / t1, t3 / t2, t4 / t3\cdots \quad (1b)$$

另外，模式指定部 8 是在于指定模式者，即指定：将上述相关运算部 4 所产生的相关数据作为声音识别处理用的匹配数据来登记在数据存储体 6 中的模式，或利用登记于数据存储体 6 中的各种匹配数据来实际进行声音识别处理的模式。

另外，当登记模式被上述模式指定部 8 所指定时，上述数据登记部 5 会将上述相关运算部 4 所产生的相关数据作为匹配数据来登记于数据存储体 6 中。

在此，数据存储体 6 为储存数据的记录媒体，在相关数据（匹配数据）的登记模式时，取入相关运算部 4 所产生的相关数据，然后予以记录。并且，在声音识别模式时，按照数据对照部 7 的要求来读出所被储存的相关数据（匹配数据），然后予以输出。

另外，数据对照部 7 会利用从相关运算部 4 输出的相关数据与从数据存储体 6 读出的匹配数据来进行图形匹配处理，由储存于数据存储体 6 的多个匹配数据中检测出来自相关运算部 4 的相关数据与在某或然率以上一致的匹配数据。由此来识别出从麦克风（未图示）等输入的声音为对应于上述检测出的匹配数据的声音，并将此结果输出至数据处理部（未图示）等。如此一来，在数据处理部中可按照所被识别出的声音内容来执行处理。

如上所述，就本实施例的声音识别装置而言，在匹配数据的登记时，先是输入人所发出的声音（模拟信号），然后再予以数字化进行运算，而由此检测出数字数据的微分绝对值为形成预定值以下的标本点。然后，产生与该标本点的振幅数据相关的相关数据，以及与表示标本点的时间间隔的定时数据相关的相关数据，且予以作为匹配数据来登记于数据存储体 6 等的记录媒体。

此外，在声音识别的处理时，同样的对人所发出的声音进行处理，而产生振幅数据的相关数据与定时数据的相关数据。然后，利用如此产生的相关数据与事先登记于数据存储体 6 中的多个匹配数据来进行图形匹配处理，以识别所被输入的声音为何。

在此，根据数据产生部 3 而产生的振幅数据与定时数据会依所输入的声音内容而形成不同的独特数据。因此，在利用由此振幅数据与定时数据所产生的相关数据进行图形匹配下，可进行所被输入声音为何等的声音识别。

并且，在本实施例中，并非是原封不动地将数据产生部 3 所产生的振幅数据与定时数据作为匹配数据使用，而是以取其比率的相关数据作为匹配数据来使用。由此而能够提高声音识别率。

即,即使是同样说话内容,也会因为此刻的发音大小及速度等的不同,其振幅数据的值与定时数据的值有所差异。因此,若原封不动地将振幅数据与定时数据作为匹配数据来使用的话,则即使是同样说话内容,也会因为此刻的发音状态不同而识别出的声音有所差异。

5        对此,如本实施例所示,取振幅数据及定时数据的比率,而以此作为匹配数据来使用,由此即使全体发音的大小及速度不同,只要说话内容相同,各标本点的振幅值的变化比率及标本点间的间距的变化比例还是会约形成相同,而使能够识别正确的声音。如此一来,例如可使用特定的话语来作为声音指令,通过声音来操作电脑或携带型终端机等。

10       此外,在上述实施例的相关运算部 4 中,虽是分别如上述式(1a)及(1b)一般来计算连续 2 个振幅数据间的比率及连续 2 个定时数据间的比率,但相关数据的计算并非只限于此例。例如,在上述式(1a),(1b)的比率计算中,分母与分子也可颠倒。并且,不只限于运算连续 2 个数据间的比率,也可运算分离的数据间的比率。

15       另外,在此虽是根据除法运算运算 2 个数据间的比率,但也可通过加法运算或减算或乘算,或者是加减乘除的任意组合运算来求得相关值。另外,并非只限于使用 2 个数据的相关运算,也可为利用更多数据的相关运算。

20       再者,上述实施例虽是分别运算振幅数据的相关值与定时数据的相关值,但也可运算振幅数据与定时数据间的相关值。

即,只要能够通过相关值的运算来使相同声音无论在何种发音状态下都形成大致相同的值即可,相关运算的方法并无特别加以限制。当然,通过相关运算方式的改变,也有可能使声音的识别率更为提高。

25       另外,即使不运算相关值,还是可以通过境界值(在利用数据对照部 7 来进行图形匹配处理时。供以判断是否为相同声音者)的调整,即通过调整与匹配数据的一致程度相关的临界值来某程度上提高声音的识别率。

另外,也可对根据相关运算部 4 所计算的相关数据再进行使下位数位化整的处理,舍去被运算的比率数据的小数点以下。另外,也可在进行该化整处理时,首先使被运算的相关数据形成数倍后再化整下位数位。

30       如此一来,即使相同的声音,还是可以无视因出声状态而具有些微差

异的相关数据的下位数位，进行图形匹配处理，提高根据相关运算部 4 而输出的相关数据与根据数据存储体 6 而读出的匹配数据的一致率，能够进而提高声音识别率。

此外，并非只是原封不动地将相关运算部 4 所取得的一连相关数据利用  
5 用于图形匹配处理，也可不将一连的相关数据中最初与最后的相关数据利用  
用于图形匹配处理。或者是在不利用数据产生部 3 所取得的一连振幅数据  
及定时数据中最初与最后的几个振幅数据及定时数据下来求得相关数据。

处理人们出声后的声音，检测在各标本点的振幅数据与定时数据时，  
要考虑在出声的最初部分与最后部分所得的振幅数据与定时数据会不正  
10 确。如此，通过去除信赖性较低的在出声最初与最后部分所产生的振幅数  
据与定时数据、或是由此作成的相关数据，然后进行类型匹配，而使能够  
更为提高声音识别率。

另外，在上述实施例的数据产生部 3，虽是从将输入模拟信号变成数  
字化数据之中，检测出含有数字数据的微分绝对值为“0”的在一定值以  
15 下的点来作为标本点，但标本点检知的方法并非限定于此。例如，由 A /  
D 转换器 2 所供给一连串的数字数据之中，比较前后位置，微分绝对值较  
小位置，即以微分绝对值的极小值所显示的位置来作为标本点来检测也  
可。

在这种情况下，作为用来检测标本点的处理，将来自 A / D 转换器 2  
20 所供给的数字数据 1 次微分后，通过将所得的微分绝对值再次微分执行 2  
重微分，将该 2 重微分值的极性由负值或是零即将变化为正值的点，作为  
标本点而抽出。再者，如此根据 2 重微分值的极性所抽出的各点之中，只  
将 1 次微分绝对值小于所定值的点作为正规标本点来检测的方式进行处理  
也可以。

25 即，通过 1 次微分所取得的微分绝对值的极小点，将该 1 次微分绝对  
值再微分后的 2 重微分值的极性，必定是由负值变化成正值。因此，求得  
数字数据的 2 重微分值，只要检测出由其极性从负值变化成正值的点（包  
含 2 重微分值为零的点），就能够正确地检测出 1 次微分绝对值的极小点。  
此时，即使相同值的极小点连续产生 2 个时，也能够以其中一方作为标本  
30 点确实地检测。另外，只要将 1 次微分绝对值较所定值还小的点作为正规

的标本点予以检测出，则不必把不要的点当作标本点而去作检测。

图 5，通过进行上述的 2 重微分来检测出标本点的构成例方块图，表示图 3 中的微分器 301 及标本点检测部 302 的构成例。

如图 5 所示，微分器 301，是具备有第 1 微分部 21、化整运算部 22、  
5 第 2 微分部 23。另外，标本点检测部 302，是具备有极性变化点检测部 24，以及临界值处理部 25。

上述第 1 微分部 21，如图 4 所构成，将由图 1 的 A/D 转换器 2 所供给的数字数据，按照各个取样点而微分，取其绝对值而输出。化整运算部 22，是在第 1 微分部 21 所运算的 1 次微分绝对值，进行下值数位的删除处理。该处理，是基于微分器 301 所算出的微分绝对值，进行检测标本点时，对于是否为标本点的判定上，用来保持余裕度而执行。  
10

例如，若是删去下位 1 个位，则可以将实际上所算出的微分绝对值为 0—1 的范围内的所有微分绝对值当作为 0 来判定。另外，若是删去下位 2 个位，则可以将实际上所算出的微分绝对值为 0—3 的范围内的所有微分绝对值当作为 0 来判定。在此，是利用将上述 1 次微分绝对值以 8 进行除法运算而进行下位 3 个位的删除处理。通过如此处理，可以避免受到杂音等的微小变动而有影响，而能够使不要的点不会作为标本点而检测。  
15

由上述化整运算部 22 所输出的数据，供给至第 2 微分部 23 及标本点检测部 302 内的临界值处理部 25。该第 2 微分部 23 也如图 4 的构成，更将由上述化整运算部 22 进行化整运算后的 1 次微分绝对值，在每一个取样点予以微分。由该第 2 微分部 23 所求得的 2 重微分值及表示其极性的借位 (borrow) 值，供给至标本点检测部 302 内的极性变化点检测部 24。  
20

上述极性变化点检测部 24，是由微分器 301 内的第 2 微分部 23 所供给的 2 重微分值的极性，刚要从负值变化成正值的点，例如，极性负值的 2 重微分值在连续取得时，将最后为负值点、或 2 重微分值成为零的点作为标本点的候补而抽出。在极性无法取得连续呈负值的 2 重微分值时，则更进一步地将该负值的点作为标本点的候补而抽出也可以。  
25

然而，对于标本点的检测时，为了保有余裕度而在化整运算部 22 将微分绝对值的下位数的位舍去时，可预料得到连续 2 个同值的极小值的情况会增多。即使在此情况下，求得 2 重微分值，只要检测出其极性由负值  
30



变化成正值的点（包含 2 重微分值为零的点），就能够以同值的极小点之中的一方，作为标本点来确实地检测出。

另外，连续产生 2 个同值的极小值时，如图 4 所示，也可以根据由节点 d 所输出的差分器 31 的借位值，判断微分值的正负值的极性，将微分值的极性变化侧作为标本点而检测。

另外，微分值的极性没有变化时，也可以参考位在 2 个同值的连续取样点前后的取样点的微分绝对值大小关系，以其值靠近较小一方的点作为标本点而检测。

临界值处理部 25，对于由上述极性变化点检测部 24 所抽出的标本点的候补，比较由化整运算部 22 所供给的 1 次微分绝对值与预先订定的临界值，仅将 1 次微分绝对值比临界值还小的点作为正规的标本点来检测，然后传达至图 3 的定时产生部 303。

另外，在上述图 5 的例中，在化整运算部 22，虽是以实施化整运算后的 1 次微分绝对值进行临界值处理，但也可以使用在第 1 微分部 21 所求得的，实施化整运算前的 1 次微分绝对值进行临界值处理。另外，上述图 5 中，虽然以 2 重微分值的极性由负值纲要变化成正值的点来作为标本点而抽出，但也可以使用负值刚变化成正值后的点来作为标本点而抽出。

另外，作为检测标本点的其他方法，可以不使用微分绝对值，而根据含有极性的微分值检测标本点。例如，将由 A / D 转换器 2 所输出的数字数据进行 1 次微分，以其微分值的极性变化的点作为标本点来检测也是可能的。此时，图 3 的微分器 301，便将来自 A / D 转换器 2 所输入的数字数据予以 1 次微分。此时微分器 301，在每次接受到一定频率的输入时钟 CK0 时，便对数字数据进行微分。微分值，是通过在某输入时钟 CK0 的时间下，由所取得当前数据减去时间上之前 1 个数据而求得。此时，没有数据处为预定值。

另外，标本点检测部 302，除了输入有由微分器 301 所计算出的微分值之外，也输入有由 A / D 转换器的所输出的数字数据。并且，根据这些数据，将数字数据的微分值的极性变化的点作为标本点而检测出。

例如，标本点检测部 302，首先，检测出微分值的极性由正变为负的点、微分值的极性由负变为正的点、以及微分值为 0 的点。然后，对于微

分值的极性由正变为负值的点，其正负两侧的点之中，将来自 A / D 转换器 2 的数字数据值为大的一方，作为标本点而检测出。另外，微分值的极性由负变为正的点，其正负两侧的点之中，以来自 A / D 转换器 2 的数字数据值较小的一方作为标本点而检测出。另外，微分值为 0 的点，则将该点本身作为标本点来检测。另外，微分值为 0 的点连续有 2 个以上出现时，例如，将其两端位置作为标本点来检测。

定时产生部 303，是计算从检测出 1 个标本点至检测上下一个标本点为止所供给的时钟 CK0 的数目，将其作为定时数据输出的同时，输出代表各个标本点的检测点定时的定时时钟。

另外，上述振幅产生部 304，是按照该定时产生部 303 所输出的定时时钟，仅取出该标本点位置的数字数据作为振幅数据来输出。

其次，说明利用本实施例的声音合成装置。

在本实施例的声音合成装置中，是将表示特定的单字或文章等文本数据，以及由对应该单词或文章的声音信号，通过在图 1 的声音识别装置，至数据产生部 3 为止的处理，将所产生的振幅数据及定时数据的组合，相互联系。

然后，由所指定的文本数据将声音合成而输出时，使用在该文本数据被相联系的振幅数据与定时数据，通过进行如后述的插入运算，插入于各个振幅数据之间而产生插入数据，然后将其 D / A 转换后输出。

图 6，是利用本实施例，表示声音合成装置的构成例的方块图。如图 6 所示，本实施例的声音合成装置，是由：文本解析部 11、数据读出部 12、数据存储体 13、定时发生器 14、D 型触发器 15、插入处理部 16、D / A 转换器 17、及 LPF18 所构成。

数据存储体 13，是将表示各种音节、单字或文章的文本数据所对应的声音而产生的振幅数据及定时数据的组合，与上述文本数据相联系而存储。在此被存储的振幅数据及定时数据，是通过与图 1 的声音识别装置的数据产生部 3 为止施以相同处理所产生的。

另外，在此，由对应文本数据的声音所产生的振幅数据及定时数据，虽然与其文本数据相联系而存储，但也可以由图示、CG 数据、图像数据等文本以外形式的数据所对应的声音（例如，解说声音等）所产生的振幅

数据及定时数据予以对该数据相联系来存储也可以。

文本解析部 11, 解析所指定需要的文本数据, 识别其音节、单字或文章等的内容。另外, 图示、CG 数据、图像数据等被指定时, 则解析这些并识别所指定为何。数据读出部 12, 是按照文本解析部 11 的解析结果, 5 从数据存储体 13 内读出对应所指定的文本数据等内容的振幅数据与定时数据。

定时发生器 14, 接受到从数据存储体 13 所读出的定时数据之后, 从所定频率的输入时钟 CK0 产生表示以该定时数据所示的不定时间间隔的读出时钟。D 型触发器 15, 是将存储在数据存储体 13 中, 将与上述定时 10 数据成组合的振幅数据, 依由上述定时发生器 14 所产生读出时钟的定时, 顺序取入而保持, 再的输出至插入处理部 16。

在该插入处理部 16, 被输入有上述 D 型触发器 15 的输入输出段的振幅数据, 也就是被输入有在某读出时钟的时间点, 被保持在 D 型触发器 15 的振幅数据, 以及在下一个读出时钟的时间点, 应被保持在 D 型触发器 15 15 的振幅数据 (在连续 2 个标本点的 2 个振幅数据)。

插入处理部 16, 使用如此被输入的 2 个振幅数据, 以及由定时发生器 14 所输入的定时数据, 通过后述的插入运算或是卷积运算等, 在各标本点之间产生数位插入数据。然后, 将以如此方式所产生的数位插入数据通过 D/A 转换器 17 转换成模拟信号之后, 经由 LPF18 作扇模拟的合成 20 声音信号而输出。

其次, 以图 7 说明在上述插入处理部 16 的数据插入处理的原理。当以与图 1 的数据产生部 3 相同样的处理, 施于图 2 所示的声音信号而产生的振幅数据与定时数据的组合被存储在数据存储体 13 中时, 其数据呈 (※, 7) (5, 3) (7, 9) (3, 1) (3, 6) (3, 3) 的数值列。另外, ※ 25 是在图 2 中表示没有图示的数值。从数据存储体 13, 根据在此所示的顺序读出数据。

即, 首先最初由数据存储体 13 所读出的振幅数据值 “Z” 与定时数据值 “5” 的 2 个数据值, 由插入运算产生波形 a1 的数据。其次, 由上述的定时值 “5”, 以及接着被读出的振幅数据值 “3” 的 2 个数据值, 由插入 30 运算波产生形 a2 的数据。

其次，由上述的振幅数据值“3”，以及接着被读出的定时数据值“7”的2个数据值，由插入运算产生波形b2的数据。再有，由上述的定时数据值“7”，以及紧接着被读出的振幅数据值“9”，由插入运算产生波形b1的数据。以下同样地，顺序读出振幅数据值与定时数据值的组合，依次产生波形c1、c2、d2、d1、e1、e2的数据。

通过以上的处理，产生波形a1、b1、c1、d1、e1连续性的数位信号（图7的上段），以及波形a2、b2、c2、d2、e2连续性的数位信号（图7的下段）。然后，将如此方式所产生的2个数位信号相互加法运算，通过数位/模拟转换，合成为具有图2所示波形的模拟声音信号。

图8为表示取出图7所示的时刻T1—T2的区间的图，图8(a)为加法运算前的2个波形a1、a2，图8(b)为通过加法运算而产生的合成波形 $a1+a2$ 。

对于图8(a)，D1为在时刻T1的振幅数据值（在图7的例为“7”），D2为在时刻T2的振幅数据值（在图7的例为“3”），T为时刻T1—T2间的时间间隔的定时数据值（在图7的例为“5”），t为时刻T1—T2间的任意的时间点。

如图8(a)所示，使用在时刻T1的振幅数据值D1，以及表示时刻T1—T2间的时间间隔的定时数据值T，以时刻T1—T2间的任意时间t为变数，即某随着根据取样频率的时钟CK0，将时间t的值1个个增加，通过插入运算产生波形a1的数据。

另外，使用在时刻T2的振幅数据值D2，以及表示时刻T1—T2间的时间间隔的定时数据值T，同样地，以时间t为变数，通过插入运算产生波形a2的数据。

然后，以如此所产生的波形a1、a2的数据作为上述时间t的变数，通过加法运算，合成如图8(b)的波形。

以下，说明利用如上述的插入处理而能够合成声音信号的原理。

一般而言，为了由离散的数字数据连续性地取得模拟信号，将以离散性所输入的数字数据之间予以插入，提高拟似性取样频率。通常，如此的数据插入，是使用一定的标本化函数来进行。

在图9表示以sinc函数来作为标本化函数的一例。在图9所示的sinc

函数的例中，只有  $t=0$  的取样点的值为“1”，等间隔的其他全部取样点（ $t=\pm 1、\pm 2、\pm 3、\pm 4、\dots$ ）的值全部为“0”。

图 10 是用以说明使用如此标本化函数的一般的数据插入动作的图。在图 10 中，以等间隔的取样点  $t_1、t_2、t_3、t_4$  在各别的离散数据值作为  $Y(t_1)、Y(t_2)、Y(t_3)、Y(t_4)$ ，例如，在考虑取样点  $t_2$  与  $t_3$  之间，求得一定位置  $t_0$ （由  $t_2$  的距离  $a$ ）所对应的插入值  $y$  时。

一般，使用标本化函数来求得插入值  $y$  时，是对于所给予的各别离散数据求得在插入位置  $t_0$  的标本化函数的值，以之进行卷积运算就可以。具体而言，是使得在  $t_1-t_4$  的每一个取样点，在标本化函数的中心位置的峰值高度一致，求得此时各别在插入位置  $t_0$  的标本化函数的值（以  $\times$  记号表示），再将其全部加法运算。

如此地将插入处理，以随着时间经过（随着取样时钟增加），使插入位置  $t_0$  依次移动，按顺序求得连续变化的插入值  $y(t_0)$ 。由此，得以能够取得使各离散数据间，更加平滑连接的数位波形，将其通过 D/A 转换，而可以取得连续的模拟信号。

本实施例应用如此方式的数据插入处理。即，如图 8(a) 所示，在时刻  $T_2$ ，由已输入的振幅数据值  $D_1(=7)$  与定时数据值  $T(=5)$ ，在第 1 标本点（时刻  $T_1$ ）、求得形成取“0”以外的值的标本化函数的一部分的波形  $a_1$ ，同时由振幅数据值  $D_2(=3)$  与定时数据值  $T(=5)$ ，在第 2 标本点（时刻  $T_2$ ），求得形成取“0”以外的值的标本化函数的一部分的波形  $a_2$ 。

并且，使这些波形  $a_1、a_2$  的值，通过个别加法运算上随着时间经过同时依次移动的插入位置  $t$ ，来取得离散数据  $D_1、D_2$  间，更加平滑连接的数位波形，将其通过 D/A 转换，取得连续的模拟信号。

然而，在本实施例如图 2 所示，是将数字数据使其微分绝对值，以在一定值以下的点的时间间隔标本化而取得离散数据。因此，取得离散数据的各个标本点的间隔，并不仅限于要保持等间隔，较多的情况是呈不定的间隔（即使在图 2 的例中，各标本点的间隔不定为“5、7、3、3、3”）。

在此，在声音合成装置，例如求得时刻  $T_1-T_2$  间的插入值时，如图 8，以该时刻  $T_1、T_2$  的标本点间的时间间隔，仅使用标本化函数  $a_1、$

a2, 进行上述的卷积运算, 对于标本点间的时间间隔相异于此的其他标本化函数 b1、b2、c1、c2、d1、d2、e1、e2, 并在该卷积运算时并无须考虑。

另外, 例如在求得时刻 T2—T3 间的插入值时, 以其时刻 T2、T3 的标本点间的时间间隔 (=7), 仅使用标本化函数 b1、b2 来进行卷积运算, 对于标本点间的时间间隔不同于此的标本化函数 a1、a2、c1、c2、d1、d2、e1、e2, 并没有考虑关于该卷积运算时的考虑。求得其他标本点间的插入值时也相同。

其次, 对于上述数据插入的具体处理例说明如下。如上所述, 例如在求得时刻 T1—T2 间的插入值时, 仅使用由在时刻 T1、T2 的各个振幅数据值, 以及表示时刻 T1—T2 间的时间间隔的定时数据值所求得的标本化函数 a1、a2。即, 用以求得在时刻 T1—T2 间的各插入位置 t 上的插入值时的必要数据, 全部皆可在时刻 T2 的时点取得, 在此时点如图 8 (b) 所示, 能够合成信号。

在此, 在本实施例中, 每取得按 T1—T6 的各个离散时刻的 2 个振幅数据值 D1、D2, 以及表示其时间间隔的定时数据值 T, 使用这些数据值, 按照以下所说明的插入计算式, 通过算出插入值, 使数位波形依次合成。图 11 是用以说明该插入运算式的说明图。

如图 11 所示, 具有振幅数据值 D1、D2 的 2 个标本点间的插捕值, 是能够使插入位置 t 相关的 2 个 2 次函数 x1、x2, 通过刚好在中间时点连续性的函数来表示。即, 在本实施例中, 将 2 个标本点之间分成前半部与后半部, 分别使用 2 次函数 x1、x2 来计算插入值。

在此, 为标本点间的时间间隔的定时数据值 T, 有时是奇数也有时是偶数, 在为奇数时, 会产生插入位置 t 不会到正中间时点上的状态。因此, 振幅数据与定时数据产生时, 通过执行 2 倍的过取样, 使所取得的定时数据值一直为偶数的方式也可以。此情况下, 图 2 所示的 5 个定时数据值“5、7、3、3、3”, 通过实际上的倍的过取样而成“10、14、6、6、6”的值而存储在数据存储体 13。在图 11 是表示由过取样后的 2T 的标本点间的时间间隔。

在图 11 中, 2 个 2 次函数 x1、x2 分别表示为

$$x1 = D1 + at^2 \quad \dots (1)$$

$$x2 = D2 - a(t - 2T)^2 \quad \dots (2)$$

另外，这些函数  $x1$ 、 $x2$ ，由于在连续的标本点的正中间时点下为连续，所以

$$5 \quad x1 = x2 \quad (t = T) \quad \dots (3)$$

在此，将式 (1) (2) 代入式 (3)，则

$$D1 + aT = D2 - aT^2 \quad \dots (4)$$

解出  $a$  为

$$a = -(D1 - D2) / 2T^2 \quad \dots (5)$$

10 因此，通过将式 (1) (2) 代入式 (5)，可得到

$$x1 = D1 - \{(D1 - D2) / 2T^2\} t^2 \quad \dots (6)$$

$$x2 = D2 + \{(D1 - D2) / 2T^2\} (2T - t)^2 \quad \dots (7)$$

即，按照 2 倍于原本取样频率的时钟，以依次增加的插入位置  $t$  作为变量，进行上述式 (6) (7) 的运算，便能够将原本的离散数据间更加平滑地连接而合成数位波形。在本实施例中，如此的插入运算处理，是按照来自振幅数据值与定时数据值所成的信号列，分别输入于各个离散时刻  $T1$ — $T6$  而按顺序进行。

即，在图 7 的例示，在时刻  $T1$ 、 $T2$  的标本点的振幅数据值与其间的定时数据值被输入后的时点，进行其标本点间的插入运算，立即合成平滑的数位波形。再有，在时刻  $T3$  的标本点的振幅数据值与标本点  $T2$ — $T3$  间的定时数据值被输入后的时点，进行该两者之间的插入运算，立即合成平滑的数位波形。以下相同样地依次进行处理。

如以上所说明，在本实施例中，在来自实际声音所产生的各标本点，先将表示振幅数据与各标本点间的间隔的定时数据与文本数据相联系、当所期望的文本数据被指定时，由与其相对应的振幅数据与定时数据，通过插入运算，合成模拟声音信号后输出。由此，与利用事先录音后的声音信号相比，将声音合成（再生）的方式，只要先存储极少量的数据（振幅数据及定时数据），就能够由文本数据等合成出接近真人声音的高质量声音。另外，由于可以将所读出的数据通过简单的插入运算依次予以处理合成声音，所以可以实现即时（real time）动作。

另外，上述式（6）（7）所示的插入运算处理，可以通过逻辑电路等的硬件构成来实现，也可以通过 DSP（Digital Signal Processor）或是软件（储存在 ROM 或 RAM 的程序）来实现。

## 5 实施例 2

以下说明本发明的实施例 2。

在实施例 2 中，声音识别装置，至少在由数据产生部 3 产生振幅数据与定时数据前的阶段，对于所给与的数字数据，利用进行  $n$  倍过取样与移动平均运算或是卷积运算（以下，称之为卷积（convolution）运算），通过  
10 将离散性数据之间予以插入，来取得更平滑的数据。

图 12 是按实施例 2 的声音识别装置全体构成例的方块图。如图 12 所示，实施例 2 的声音识别装置，相对于在图 1 所示的实施例 1 的声音识别装置，追加有过取样电路 9 与 PLL（Phase Locked Loop）电路 10。

上述过取样电路 9，位于 A/D 转换器 2 与数据产生部 3 之间，对由  
15 A/D 转换器 2 所输入的数字数据，通过进行  $n$  倍的过取样与卷积运算，来求得填补离散数据之间的数位的插入值。例如，过取样电路 9，以 44.1KHz 的频率输入被取样的声音数据，将其以 8 倍频率（352.8KHz）进行过取样的同时，并执行卷积运算。然后，将所得到一连串过取样数据输入数据产生部 3。

20 数据产生部 3，由上述过取样电路 9 所供给的一连串过取样数据之中，通过上述任一方法检测出标本点。然后，在测得的各标本点的振幅数据值，与代表显示各标本点的时间间隔的定时数据值的组合，因应由模式指定部 8 所指定该时点的模式，输出至数据登记部 5 或数据对照部 7。

另外，PLL 电路 10，由基准频率（例如 44.1KHz）的输入时钟 CK0  
25 产生 8 倍频率（352.8KHz）的时钟 CK1，并将的供给至上述过取样电路 9 及数据产生部 3。在过取样电路 9 及数据产生部 3，乃与该 8 倍频率的时钟 CK1 同期而动作。

另一方面，在声音合成装置，若要将根据 8 倍频率的时钟 CK1 所产生的振幅数据与定时数据数据存储在存储体 13 时，则对于在图 6 所示的  
30 实施例 1 的声音合成装置，必须追加没有图示出的时钟发生器。



该时钟发生器，是由基准频率的输入时钟 CK0 产生 8 倍频率的时钟 CK1，然后将的供给至定时发生器 14、插入处理部 16 及 D / A 转换部 17。

其次，对于上述图 12 所示的声音识别装置内的过取样电路 9 的构成及动作，详细说明如下。

5        在本实施例的过取样电路 9，在求得 2 个离散数据间的补间值时，是分别对包含该 2 个离散数据的 n 个离散数据的值，具有因应该值的振幅的基本波形的数字数据，进行过取样的同时，将所取得的 n 个数据相互之间，通过卷积运算而合成，以求出填补上述 2 个离散数据间的插入值的数位性值。

10       图 13 是在本实施例使用的数位基本波形的说明图。在图 13 所示的数位基本波形是由过取样进行数据插入时使用的标本化函数的基本。该数位基本波形，按每一基准频率的 1 时钟（CK0）使数据值变化于-1、1、8、8、1、-1 而作成。

在以下，使用图 14 举出如图 13 所示的归一化的数位基本波形，由离散数据值（-1、1、8、8、1、-1）/ 8，通过 n 倍的过取样与卷积运算产生插入值的情况为例，说明以本实施例的数据插入的原理。另外，在图 14 中，由于图面关系，表示以实施 4 倍的过取样为例，但图 12 的过取样电路 9，是实施以实际的 8 倍过取样。

在图 14 中，最左列所表示的一连串数值列，是对原来的离散数据值 20       （-1、1、8、8、1、-1）/ 8，进行 4 倍过取样之值。另外，从最左朝右 4 列分的数值列，是以最左列所示的数值列向下方各下降 1 格而位移。图 14 的列方向是表示时间轴，数值列朝下方向位移的含意，即是对应由最左列所示的数值列慢慢地迟延。

即，由左图 2 列的数值列，是对最左列所示的数值列，为以仅偏差 4 25       倍频率的时钟 4CLK 的 1 / 4 相位量的数值列。另外，由左起第 3 列的数值列，是对由左起第 2 列所示的数值列，仅偏差 4 倍频率的时钟 4CLK 的 1 / 4 相位量的数值列，由左起第 4 列的数值列，是对左起第 3 列所示的数值列，仅更偏差 4 倍频率的时钟 4CLK 的 1 / 4 相位量的数值列。

另外，左起第 5 列的数值列，是将对应第 1—4 列的各数值列的行相 30       加后的值除以 4 的值。通过由该左起第 5 列为止的处理，随着 4 相的卷积

运算，使 4 倍的过取样以数位性地实行。

从上述第 5 列朝右 4 列分的数值列（由左起 5—8 列的数值列），是将第 5 列所示的数值列，1 个个向下方位移。另外，左起第 9 列的数值列，是将第 5—8 列的各数值列在对应的行相互相加后除以 4 的值。通过该左起第 9 列为止的处理，并将随着 4 相的卷积运算的 4 倍过取样以数位方式执行 2 次。

另外，左起第 10 列的数值列，是将第 9 列所示的数值列向下方位移 1 行。另外，左起第 11 列（最右侧列）的数值列，是将第 9 列的数值列与第 10 列的数值可在对应的行相互相加后除以 2 的值。使该最右侧的数值列为求得目的的插入值。

使该图 14 的最右列所示的最终所得的数值列予以图示化的是图 15。具有如图 15 所示的波形的函数，是可在全区域 1 次微分，沿着横轴的标本位置  $t$  从 1 到 33 的区间时具有 0 以外的有限值，而在其以外区域的值全部为 0 的函数。

另外，函数的值在局部性区域具有“0”以外的有限值，在其以外的区域为“0”的情况称为“有限台”。

另外，图 15 的函数，具有仅在  $t=17$  的标本点取极大值，而在  $t=1$ 、9、25、33 的 4 个标本点的值为 0 的特征的标本化函数，为了取得平滑的波形数据，需要全部通过必要的取样点。

如此，在图 15 所示的函数为标本化函数，在全区域可 1 次微分，而且在标本位置  $t=1$ 、33 时收敛为 0 的有限台函数。因此，使用图 15 的标本化函数，通过根据各离散数据进行重叠，便能够使用可 1 次微分的函数将离散数据间的值予以插入。

对于数据插入处理，一般所使用的如图 9 的 sinc 函数，由于是在  $t=\pm\infty$  的标本点收敛为 0 的函数，所以当要正确地求得插入值时，必须对应  $t=\pm\infty$  为止的各离散数据计算在插入位置的 sinc 函数的值，以该值来进行卷积运算。相对于此，在本实施例所使用的图 15 的标本化函数，由于在  $t=1$ 、33 的标本点收敛为 0，所以仅将  $t=1$ —33 的范围内的离散数据纳入考虑即可。

因此，要求得某 1 插入值时，只要考虑有限的  $n$  个离散数据的值即可，

而能够大幅削减处理量。而且，对于  $t=1-33$  的范围外的各离散数据，并非是原本应该考虑而考虑到处理量或精度等其忽略掉，而是由于理论上不必要考虑，所以不会产生舍位误差。因此，只要使用本实施例的数据插入方法，便可以正确地得到插入值，通过使用该正确的插入值进行声音识别处理，因而能够提高识别率。

图 16 是图 12 所示的过取样电路 9 的构成例的方块图。如图 16 所示，本实施例的过取样电路 9，是由具备：归一化数据存储部 41、及相位位移部 42、及多个数位乘法运算器 43a—43d、及多个数字加法运算器 44a—44c 所构成。另外，在图 16 中所示的 PLL 电路 10 与在图 12 中所示者相同。

上述归一化数据存储部 41，存储有图 14 的最右列所示地将归一化后的数据列偏移成 4 相。另外，在图 14 中，虽然表示对在图 13 所示的数位基本波形进行 4 倍过取样的例，但由于在图 12 的过取样电路 9 中，实施 8 倍过取样，所以该归一化数据存储部 41 中，存储有使数位基本波形成 8 倍过取样后，通过卷积运算的归一化数据列。该归一化数据存储部 41 所存储的 4 相的归一化数据，按照由 PLL 电路 10 所供给的时钟 CK0、CK1 而被读出，再分别供给至 4 个数位乘法运算器 43a—43d 的一方的输入端子。

另外，相位位移部 42，将从 A/D 转换器 2 所输入的离散数据的相位偏移成 4 相来进行相位位移处理。通过该相位位移部 42 所产生的 4 相离散数据是按照从 PLL 电路 10 所供给的时钟 CK0、CK1 而输出，再分别供给至 4 个数字乘法运算器 43a—43d 的另一方的输入端子。

上述 4 的数字乘法运算器 43a—43d，是将从上述归一化数据存储部 41 所输出的 4 相归一化数据，以及从上述相位位移部 42 所输出的 4 相离散数据分别予以乘算。在这些的后段，所连接的 3 个数位加法运算器 44a—44c，是将上述 4 个数位乘法运算器 43a—43d 的乘法运算结果全部予以加法运算，再将其加法运算结果输出至图 12 的数据产生部 3。

由该图 16 所示的构成也可明白，在本实施例中，如图 14 所示，将通过卷积运算所得到的最右列的归一化数据，先予以存储在 ROM 等的归一化数据存储部 41。然后，使该归一化数据，根据所输入的离散数据的值的振幅而变调，并将由此所得的数据，通过 4 相的卷积运算部成而输出。

将输入离散数据的振幅值，对图 13 所示的数字基本波形相乘，对于如此所得取的数据值，也可以如图 14 所示的卷积运算的方式进行声音识别，但若使过取样电路 9 以图 16 的方式来构成时，在实际的声音识别时，就无须进行图 14 的卷积运算，具有能够使识别处理高速化的优点。

5       另外，在过取样电路 9，是实施 8 倍的过取样，但并不限定于 8 倍。例如，2 倍、4 倍等均可。

如以上详细说明，在实施例 2 中的声音识别装置中，对所输入的离散数字数据，进行过取样及卷积运算来产生平滑地变化的连续性数据，从所取得的过取样数据可得到离散的振幅数据值，以及代表不定的时间间隔的  
10    定时数据值。

在此，执行过取样与卷积运算时，由数字基本波形所产生的函数，是在有限的标本位置，其值收敛为 0 的有限台的标本化函数，是可以 1 次微分的函数。因此，在求某 1 个插入值时，仅考虑有限数目的离散数据值即可，因而可以大幅削减处理量。并且，由于不会发生舍位误差的情况，  
15    所以可以得到正确的插入值，使用该插入值进行声音识别处理时，可以提高声音的识别率。

另外，通过在以上所说明的实施例 1、2 的声音识别及声音合成的办法，可以通过硬件构成、或 DSP、或软件的任一者来实现。例如通过软件实现时，本实施形态的声音识别装置及声音合成装置，实际上是由电脑的 CPU 或 MPU、RAM、ROM 等所构成的，通过被存储在 RAM 或 ROM  
20    的程序动作来实现。

因此在电脑上，为了达成上述本实施例的功能，可以将动作程序记录在例如 CD-ROM 的记录媒体，通过读进电脑就可以实现。作为记录上述程序的记录媒体，除了 CD-ROM 之外，也可以使用磁片、硬盘、磁带、光盘、非挥发性存储体等。  
25    

另外，不仅可以通过使电脑执行所供给的程序实现上述实施例的功能，也包含使其程序与在电脑动作的 OS（操作系统）或其他应用软件共同实现上述实施例的功能的情况，或是使所供给的程序处理的全部或是一部分，通过电脑的功能扩张卡、或是功能扩张单元来执行而实现上述实施例的功能的情况，凡有关该程序都包含在本发明的实施例中。  
30

另外，在上述所说明的各实施例，任意一个都只不过是针对实施本发明所揭示的具体例示而已，不可解释成通过这些实施例来限定本发明的技术性范围。即，在不脱离本发明的精神、或是其主要特征，都可以以各种形态来实施的。

- 5 本发明提供一种利用声音信号上的预定标本点的振幅数据与定时数据的前所未有的全新声音识别办法及声音合成方法、有助于声音识别率的提高，及合成音的高质量化，以及处理的简略化。

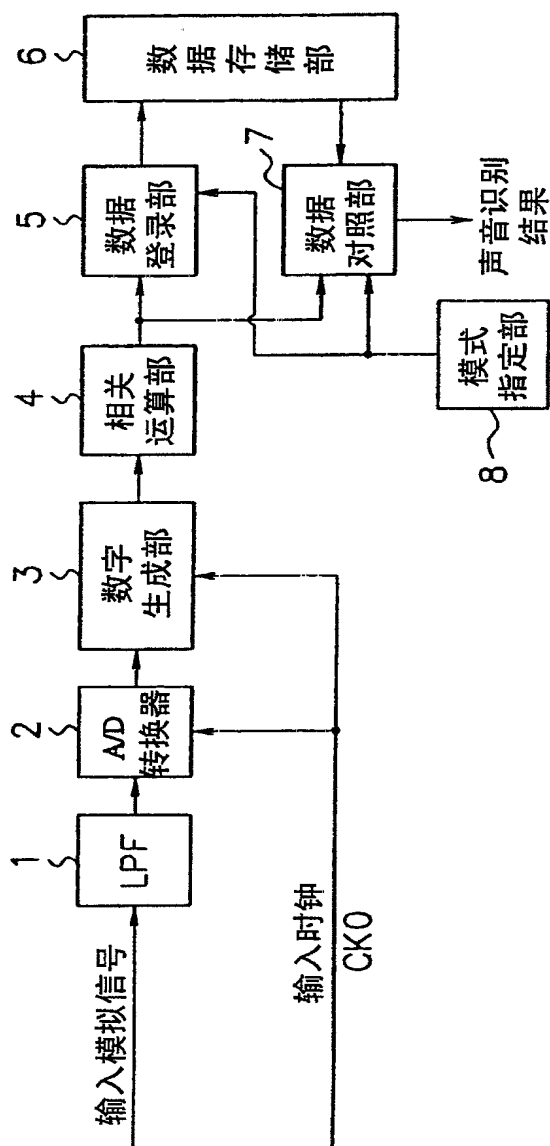


图 1

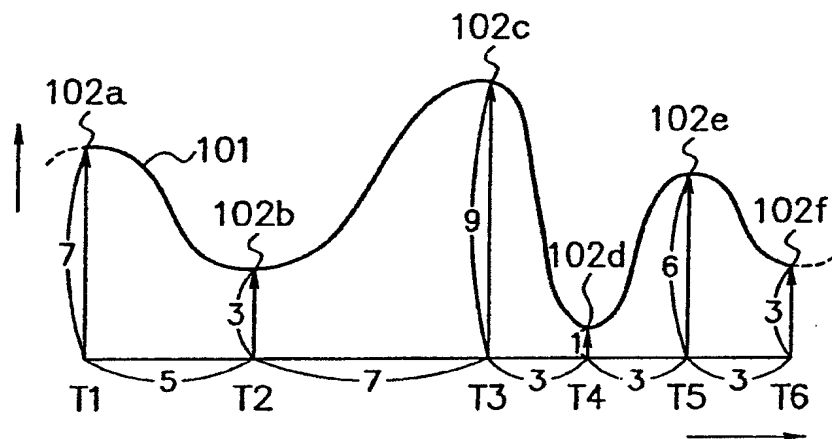


图 2

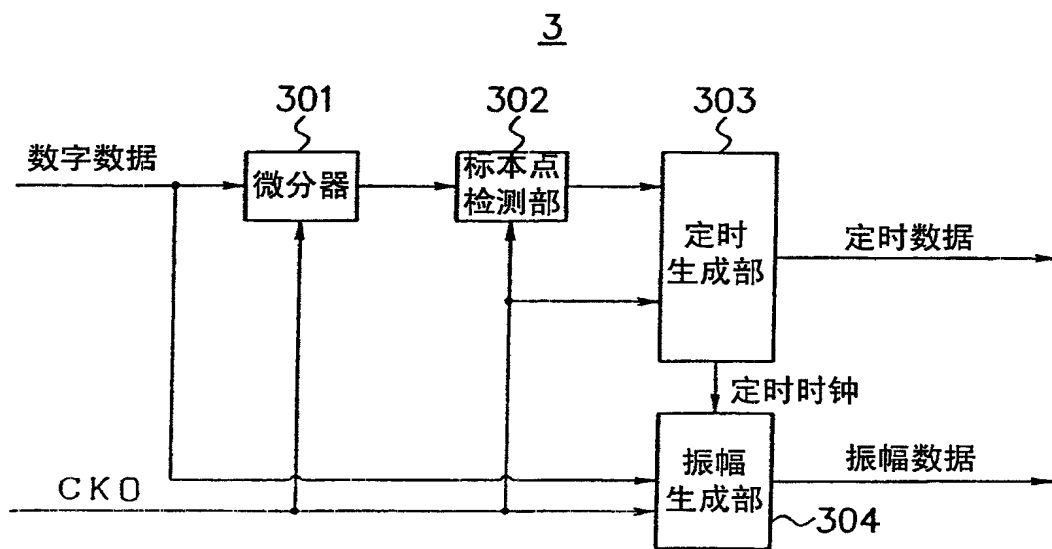


图 3

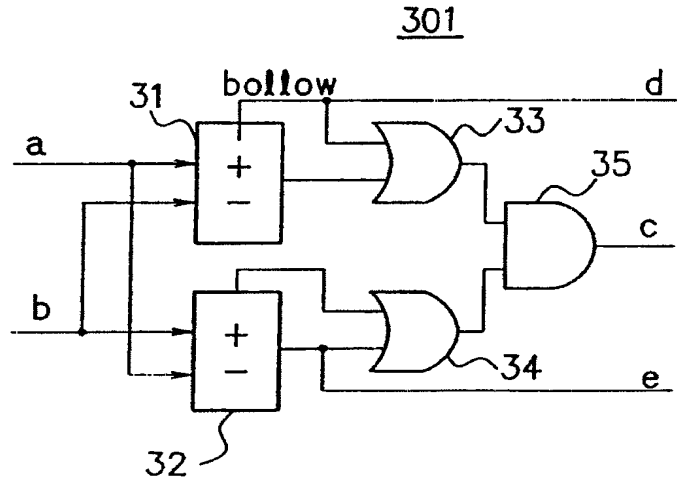


图 4

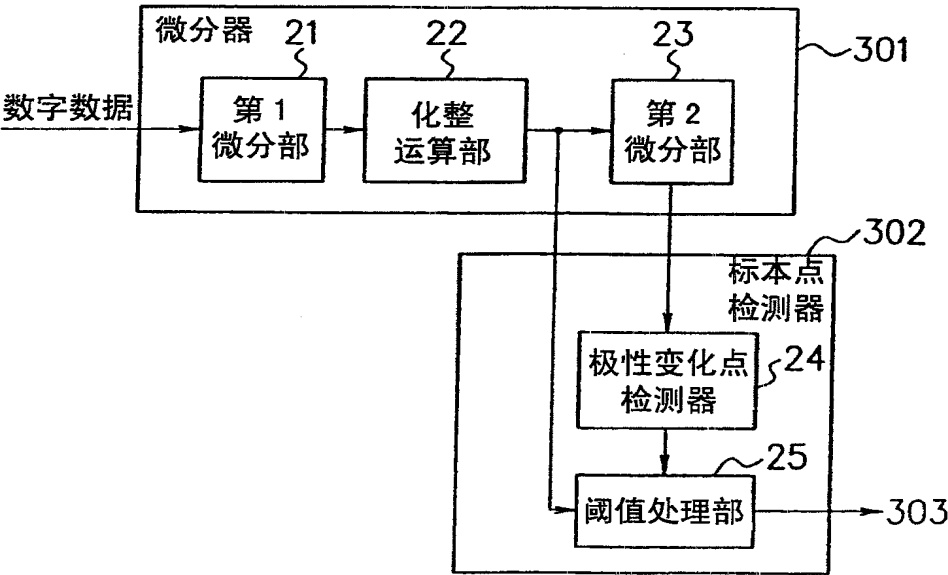


图 5



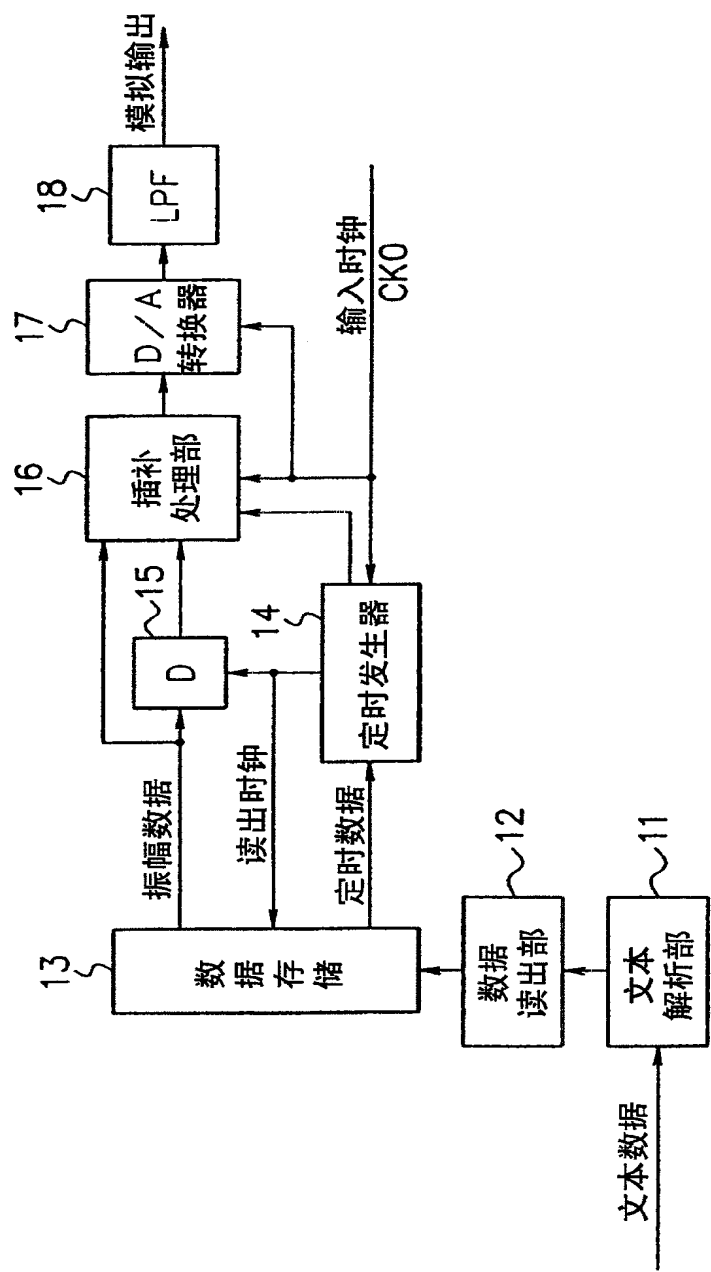


图 6

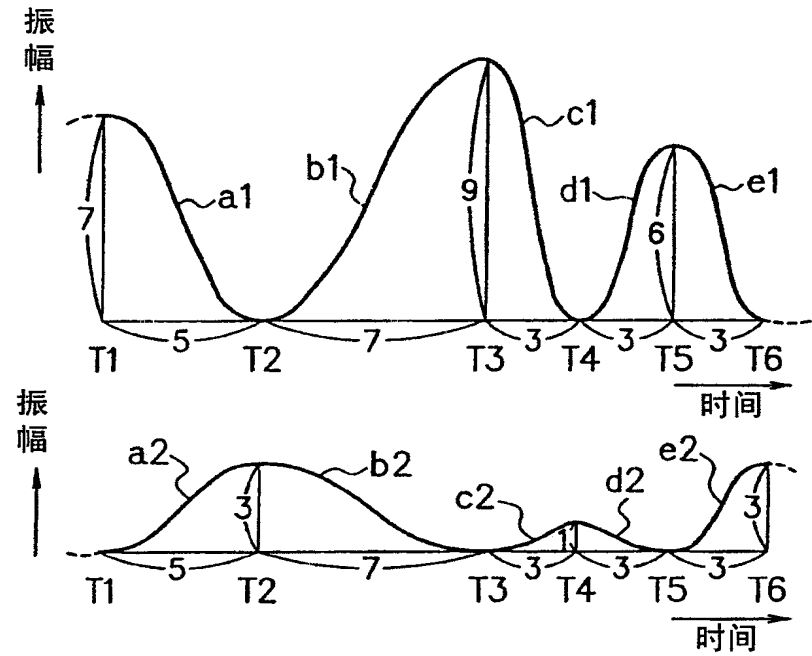


图 7

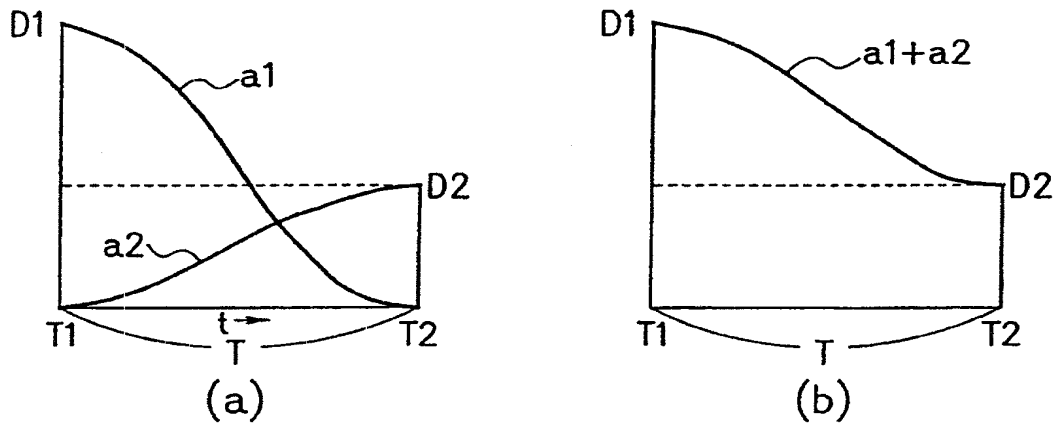


图 8

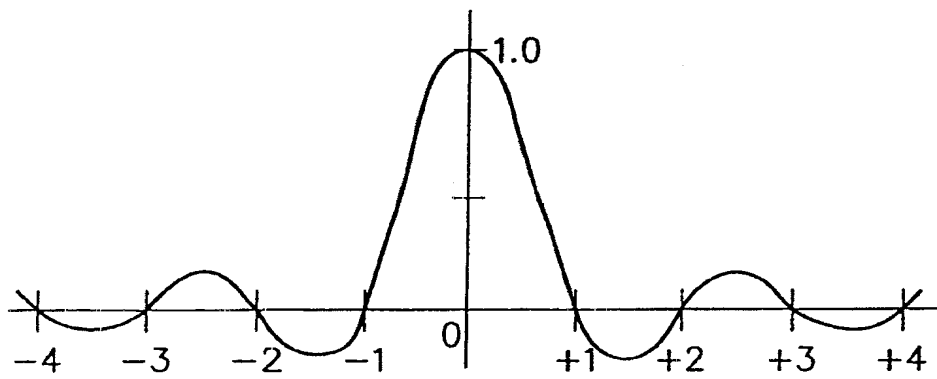


图 9

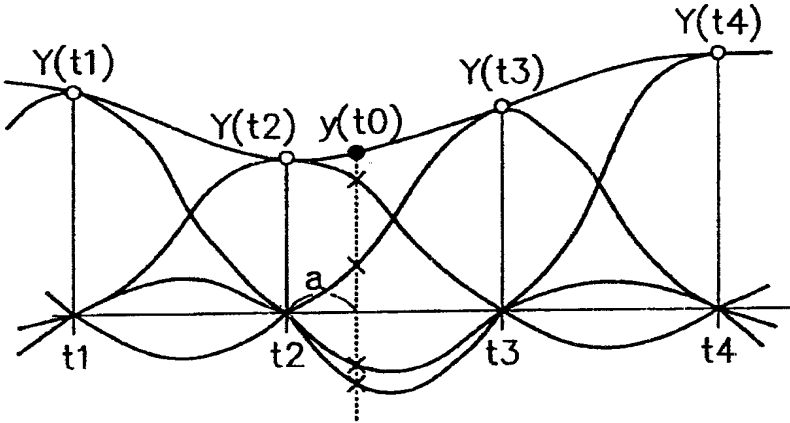


图 10

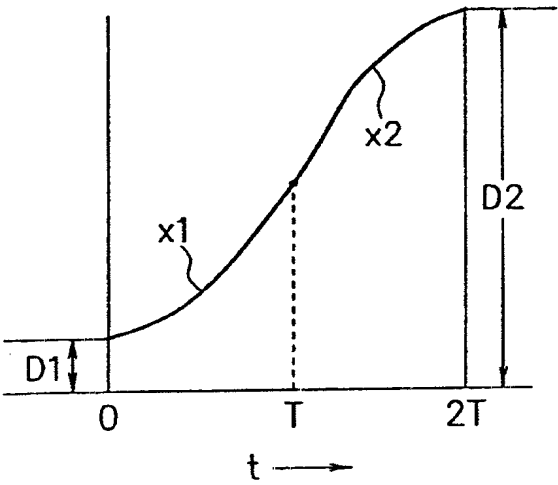


图 11

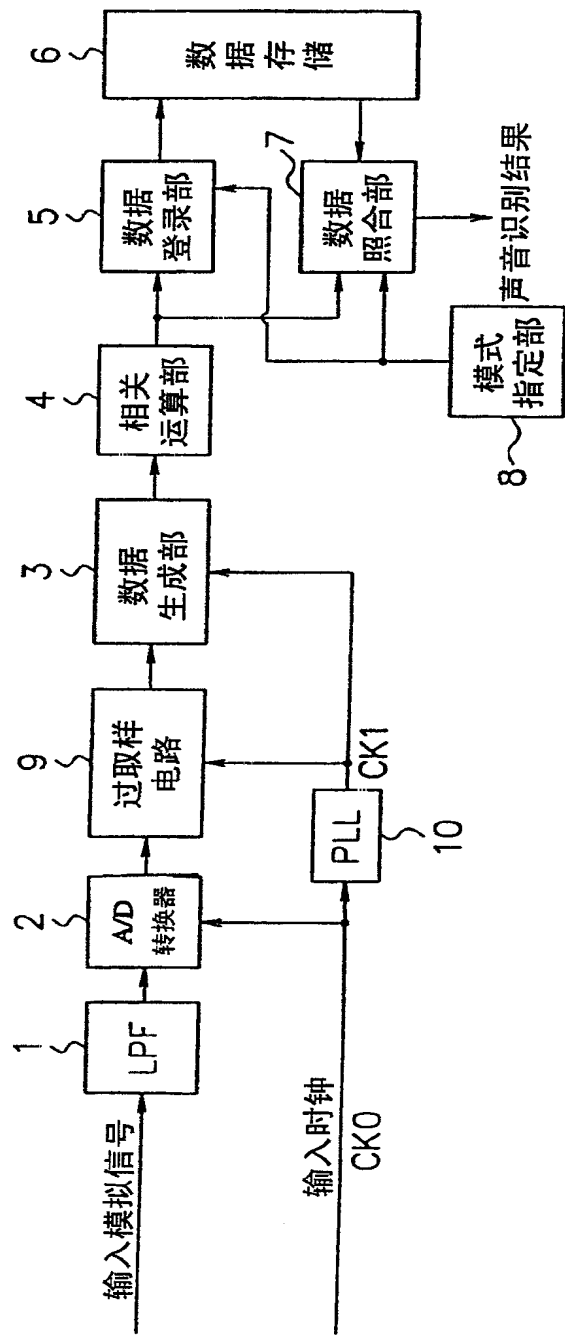


图 12

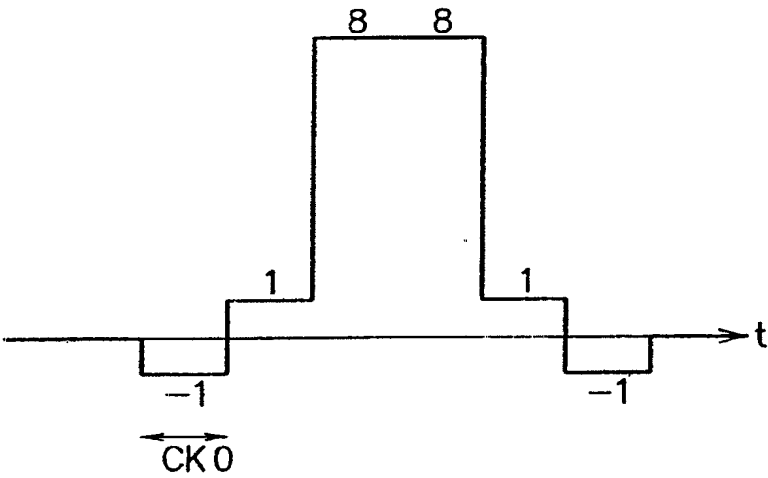


图 13

0	0	0	0	0	0	0	0	0	0	0	0
-0.125	0	0	0	-0.031	0	0	0	-0.008		-0.004	
-0.125	-0.125	0	0	-0.063	-0.031	0	0	-0.023	-0.008	-0.016	
-0.125	-0.125	-0.125	0	-0.094	-0.063	-0.031	0	-0.047	-0.023	-0.035	
-0.125	-0.125	-0.125	-0.125	-0.125	-0.094	-0.063	-0.031	-0.078	-0.047	-0.063	
0.125	-0.125	-0.125	-0.125	-0.063	-0.125	-0.094	-0.063	-0.086	-0.078	-0.082	
0.125	0.125	-0.125	-0.125	0	-0.063	-0.125	-0.094	-0.07	-0.086	-0.078	
0.125	0.125	0.125	-0.125	0.0625	0	-0.063	-0.125	-0.031	-0.07	-0.051	
0.125	0.125	0.125	0.125	0.125	0.0625	0	-0.063	0.0313	-0.031	0	
1	0.125	0.125	0.125	0.3438	0.125	0.0625	0	0.1328	0.0313	0.082	
1	1	0.125	0.125	0.5625	0.3438	0.125	0.0625	0.2734	0.1328	0.2031	
1	1	1	0.125	0.7813	0.5625	0.3438	0.125	0.4531	0.2734	0.3633	
1	1	1	1	1	0.7813	0.5625	0.3438	0.6719	0.4531	0.5625	
1	1	1	1	1	1	0.7813	0.5625	0.8359	0.6719	0.7539	
1	1	1	1	1	1	1	0.7813	0.9453	0.8359	0.8906	
1	1	1	1	1	1	1	1	1	0.9453	0.9727	
1	1	1	1	1	1	1	1	1	1	1	
0.125	1	1	1	0.7813	1	1	1	0.9453	1	0.9727	
0.125	0.125	1	1	0.5625	0.7813	1	1	0.8359	0.9453	0.8906	
0.125	0.125	0.125	1	0.3438	0.5625	0.7813	1	0.6719	0.8359	0.7539	
0.125	0.125	0.125	0.125	0.125	0.3438	0.5625	0.7813	0.4531	0.6719	0.5625	
-0.125	0.125	0.125	0.125	0.0625	0.125	0.3438	0.5625	0.2734	0.4531	0.3633	
-0.125	-0.125	0.125	0.125	0	0.0625	0.125	0.3438	0.1328	0.2734	0.2031	
-0.125	-0.125	-0.125	0.125	-0.063	0	0.0625	0.125	0.0313	0.1328	0.082	
-0.125	-0.125	-0.125	-0.125	-0.125	-0.063	0	0.0625	-0.031	0.0313	0	
0	-0.125	-0.125	-0.125	-0.094	-0.125	-0.063	0	-0.07	-0.031	-0.051	
0	0	-0.125	-0.125	-0.063	-0.094	-0.125	-0.063	-0.086	-0.07	-0.078	
0	0	0	-0.125	-0.031	-0.063	-0.094	-0.125	-0.078	-0.086	-0.082	
0	0	0	0	0	0	-0.031	-0.063	-0.047	-0.078	-0.063	
0	0	0	0	0	0	0	0	-0.031	-0.023	-0.035	
0	0	0	0	0	0	0	0	0	-0.008	-0.016	
0	0	0	0	0	0	0	0	0	0	-0.008	
0	0	0	0	0	0	0	0	0	0	0	

图 14

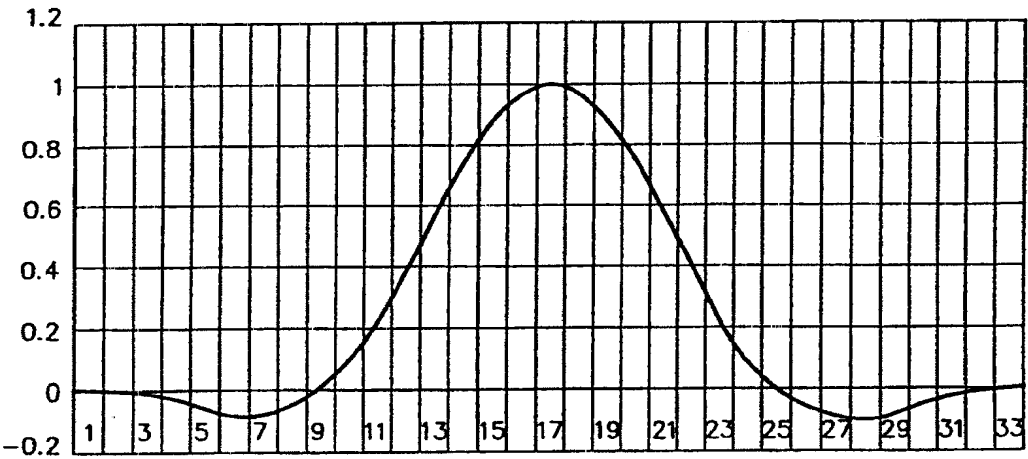


图 15



9

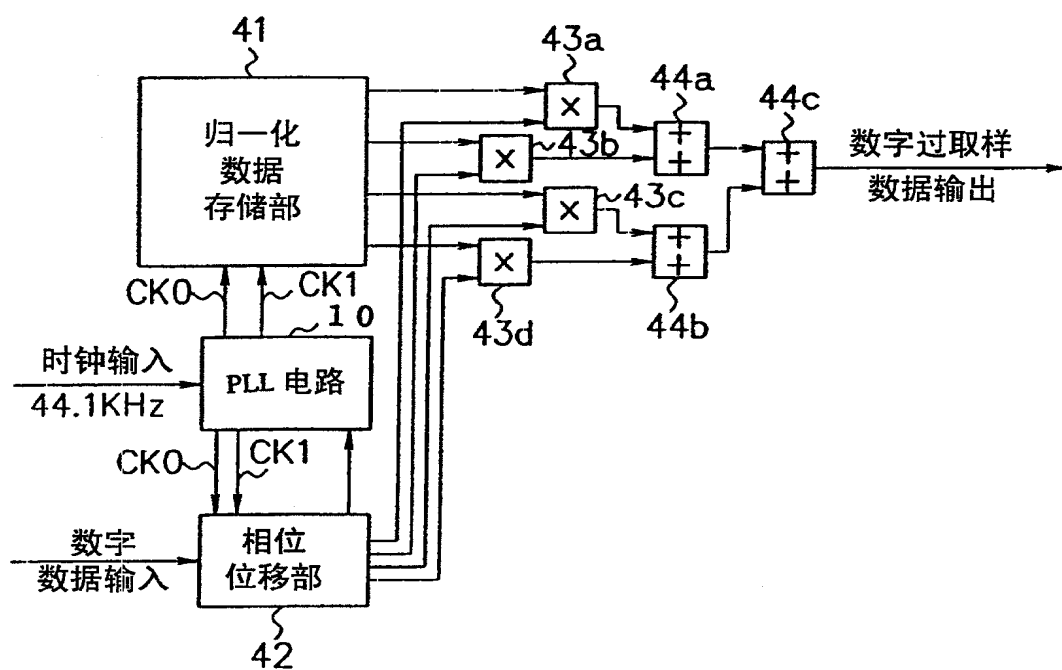


图 16